Working Paper 2000-23 / Document de travail 2000-23

The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables

by

Nikola Gradojevic and Jing Yang

Bank of Canada Banque du Canada

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2000-23

December 2000

The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables

by

Nikola Gradojevic and Jing Yang

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9

Contents

Acknowledgments

The authors are grateful to Toni Gravelle, Peter Thurlow, Angela Redish, Walter Engert, Chris D'Souza, and Nicolas Audet for their helpful comments and suggestions. We also thank Andre Bernier for providing us with data.

Abstract

Artificial neural networks (ANN) are employed for high-frequency Canada/U.S. dollar exchange rate forecasting. ANN outperform random walk and linear models in a number of recursive out-of-sample forecasts. The inclusion of a microstructure variable, order flow, substantially improves the predictive power of both the linear and non-linear models. Two criteria are applied to evaluate model performance: root-mean squared error (RMSE) and the ability to predict the direction of exchange rate moves. ANN is consistently better in RMSE than random walk and linear models for the various out-of-sample set sizes. Moreover, ANN performs better than other models in terms of percentage of correctly predicted exchange rate changes (PERC). The empirical results suggest that optimal ANN architecture is superior to random walk and any linear competing model for high-frequency exchange rate forecasting.

JEL classification: C45, F31
Bank classification: Exchange rates

Résumé

Les réseaux de neurones artificiels sont employés pour la prévision du taux de change Canada/États-Unis à fréquence élevée. Ils produisent généralement de meilleures prévisions hors échantillon récursives qu'une marche aléatoire ou un modèle linéaire. L'addition d'une variable microstructurelle (le flux des transactions) entraîne une nette amélioration de la capacité de prévision des modèles tant linéaires que non linéaires. Les auteurs font appel à deux critères pour évaluer le rendement d'un modèle : la racine carrée de l'erreur quadratique moyenne (REQM) et la capacité de prévoir le sens des variations du taux de change. La REQM est systématiquement moins élevée dans le cas des prévisions produites par les réseaux de neurones artificiels que pour celles issues d'une marche aléatoire ou de modèles linéaires, peu importe le nombre d'observations hors échantillon. De plus, les réseaux de neurones artificiels permettent de prédire correctement le sens d'un plus grand pourcentage des variations du taux de change. Selon les résultats empiriques, l'architecture optimale de ces réseaux fournit de meilleures prévisions du taux de change à fréquence élevée qu'une marche aléatoire ou tout autre modèle linéaire.

Classification JEL : C45, F31
Classification de la Banque : Taux de change

1. Introduction

Understanding exchange rate movements has long been an extremely challenging and important task. Efforts to deepen our understanding of exchange rate movements have taken a number of tacks. Initially, efforts centred on developing low-frequency fundamentally based empirical models. More recently, efforts have aimed to develop more microeconomically based models of the foreign exchange market. Throughout, model building has aimed to provide good exchange rate forecasts and improve our understanding of exchange rate movements. The models can sometimes help to pinpoint where the gaps in our knowledge lie, and suggest new avenues of research. The exchange rate forecasting model developed in this study serves all of the above purposes.

Various models aimed at explaining exchange rate fluctuations have been proposed. Meese and Rogoff (1983) found that a simple random walk model performed no worse than a range of competing representative time-series and structural exchange rate models. Out-of-sample forecasting power in those models was surprisingly low for various forecasting horizons (from 1 to 12 months).

Subsequent attempts to determine exchange rates shed little light on the problem. Baillie and McMahon (1989) pointed out that exchange rates are not linearly predictable. Similarly, Hsieh (1989) observed that exchange rate changes may be non-linearly dependent. However, Meese and Rose (1991) examined macroeconomic exchange rate models and found that the poor explanatory power of the models cannot be attributed to non-linearities. Meese and Rose (1990) used a non-parametric estimator to handle non-linearities, but this provided no significant improvement in monthly exchange rate explanation.

Each of the above-mentioned approaches tries to explain exchange rate movements with macroeconomic variables such as interest rates, money supplies, inflation rates, and trade balances. Lyons and Evans (1999) incorporated a variable reflecting the microeconomics of asset pricing into a model of the exchange rate.[1] They introduced a new variable, order flow, as the proximate determinant of the exchange rate (using daily data over a four-month period), and were able to significantly improve on existing macroeconomic models. More precisely, they managed to capture about 60 per cent of the daily exchange rate changes using a linear model.

1. His research looks at how market structure (or the trading process) has an impact on the price of the asset.

Plasmans, Verkooijen, and Daniels (1998) and Verkooijen (1996) used macroeconomic models and artificial neural networks (ANN), a very powerful tool for detecting non-linear patterns, to test whether the underlying relationship is non-linear. They could not produce satisfactory monthly forecasts. On the contrary, Zhang and Hu (1998) modelled the exchange rate as depending non-linearly on its past values, and their model outperformed simple linear models, but they never compared it to a random walk. Hu et al. (1999) showed (using daily and weekly data) that ANNs are a more robust forecasting method than a random walk model. Hence, the application of ANNs to short-term currency behaviour was successful in numerous cases and the results suggest that ANN models may offer some advantages when frequent short-term forecasts are required (Evans 1997, Jamal and Sundar 1997, Kuan and Liu 1995).[2]

This paper examines whether introducing a market microstructure variable (that is, order flow) into a set of daily observations of macroeconomic variables (interest rate, crude oil price) together with an ANN technique can explain Canada/U.S. dollar exchange rate movements better than linear and random walk models. Two statistics are used to compare models: root-mean squared error (RMSE) and the percentage of correctly predicted exchange rate changes (PERC). Empirical findings are in favour of the ANN model, which yields a very robust out-of-sample forecasting improvement in RMSE and PERC.

Section 2 describes the competing theoretical models. Section 3 introduces the ANN method and its applications to the foreign exchange (FX) market. Section 4 describes the data and the method used to assess the predictive performance of the models. Section 5 describes the empirical results of the models. Section 6 concludes the paper and recommends further research.

2. Models of Exchange Rate Determination

There are two broad theories of exchange rate modelling: traditional macroeconomic models and the more recently developed market microstructure models. Macroeconomic models estimate exchange rates at monthly or lower frequency. They usually have the following form:

$$\Delta rpfx_t = \phi\ (M_t) + \varepsilon_t,\ t=1,..,N.$$

where $\Delta rpfx_t$ is the change in the logarithm of the real exchange rate over the month or some lower frequency of observations, and M_t is a vector of typical macroeconomic variables such as the difference between home and foreign nominal interest rates, the long-run expected inflation

2. In this context, ANNs focus on daily or less-than-monthly frequency, while typical macroeconomic models are at a monthly or quarterly frequency.

differential, and relative real growth rates.[3] This paper uses a variation of the model developed by Amano and van Norden (1995):

$$\Delta rpfx_t = \varphi\ (rpfx_t, com_t, ene_t, intdiff_t) + \delta_t, t=1,..,N.$$

where $rpfx_t$ is the real Canada/U.S. exchange rate deflated by GDP deflators, com_t is the logarithm of the non-energy commodity price index (deflated by the U.S. GDP deflator), ene_t is the logarithm of the energy commodity price index (deflated by the U.S. GDP deflator), and $intdiff_t$ represents the nominal 90-day commercial paper interest rate differential (Canada–U.S.).

Macroeconomic models provide no role for any "market microstructure" effects to enter directly into the estimated equation, and thus they are incorporated through the error term δ_t.[4] These models assume that markets are efficient in the sense that information is widely available to all market participants and that all relevant and ascertainable information is already reflected in exchange rates. In other words, in this view, exchange rate changes are not informed by microstructure variables. However, typical macroeconomic models perform poorly. Moreover, empirical evidence from Lyons and Evans (1999), Yao (1997), Covrig and Melvin (1998), and this paper suggests that a microstructure variable order flow contains information relevant to exchange rate determination.

For the spot FX trader, what matters is not the data on any of the macroeconomic fundamentals, but information about demand for currencies extracted from purchases and sales orders, or order flow. It is presumed that certain FX traders observe trades that are not observable to all the other traders and, in turn, the market efficiency assumption is violated at least in the very short term.[5]

Microstructure models rely directly on information regarding the order flow. Non-dealer market participants (corporations, mutual and pension funds, etc.) analyze all the publicly available information, including macroeconomic fundamentals, and then decide on orders. Having observed order flow (which thus reflects information about macroeconomic fundamentals), dealers set their price.

---

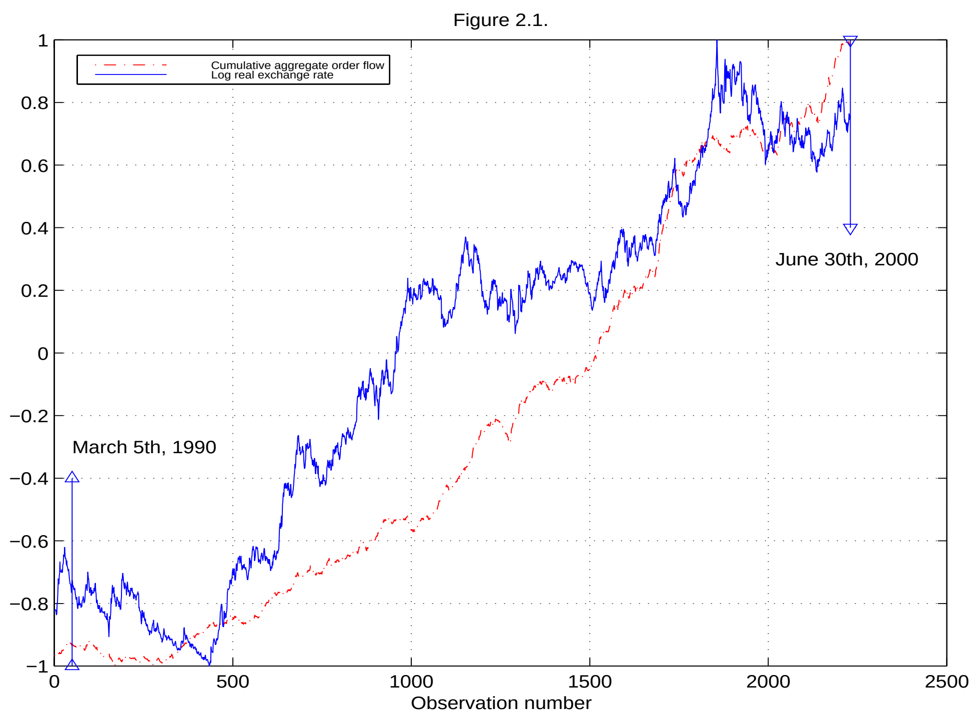
3. See Meese and Rogoff (1983).
4. Microstructure literature examines the elements of the security trading process: the arrival and dissemination of information; the generation and arrival of orders; and the market architecture, which determines how orders are transformed into trades. Prices are discovered in the marketplace by the interaction of market design and participant behaviour.
5. It may be that, without these market microstructure frictions, markets would be efficient, but trading frictions impede the instantaneous embodiment of all information into prices.

The market microstructure approach assumes the following relationship between the exchange rate and the driving variables:

$$\Delta rpfx_t = \psi\,(\Delta x_t, \Delta I_t, N_t) + \chi_t,\ t=1,..,N.$$

where Δx_t represents order flow, ΔI_t a change in net dealer positions, and N_t any other microeconomic variable. Order flow can be positive (net dollar purchases), or negative (net dollar sales).[6] Macroeconomic effects are incorporated into error term χ_t. A positive relationship between the exchange rate and order flow is expected, since informational asymmetries gradually affect the price until it reaches equilibrium. Figure 2.1 illustrates the explanatory power of an aggregate order flow (the data cover the period from January 1990 to June 2000 at a daily frequency).

Figure 2.1 Aggregate (cumulative) order flow and log Canada/U.S. real exchange rate. Note: All values are normalized to [-1,1].



As the solid line indicates, the Canadian dollar has depreciated throughout most of the sample. The relationship between the exchange rate and order flow is quite clear as a positive

6. Order flow is explained in Section 4.

correlation between cumulative purchases of U.S. dollars and the depreciation. However, it would be inappropriate to assume that order flow contains all the information that is relevant for exchange rates.

This paper combines macroeconomic and microstructure approaches into a single high-frequency data model. More specifically, it embodies modified models from Amano and van Norden (1995, 1998) and Lyons and Evans (1999):

$$\Delta rpfx_t = \Psi\ (\Delta intdiff_{t-j}, \Delta oil_{t-j}, aggof_{t-j}) + \eta_t,\ j=\{1, 7\};\ t=1,..,N.$$

where $\Delta intdiff_t$ is the change in the differential between the Canadian and U.S. nominal 90-day commercial paper rates, Δoil_t is the daily change in the logarithm of the crude oil price, and aggregate order flow is denoted by $aggof_t$. In Section 4, $aggof_t$ is disaggregated and individual order flows are considered.

ANNs are employed to estimate a non-linear relationship between exchange rate movements and these variables.

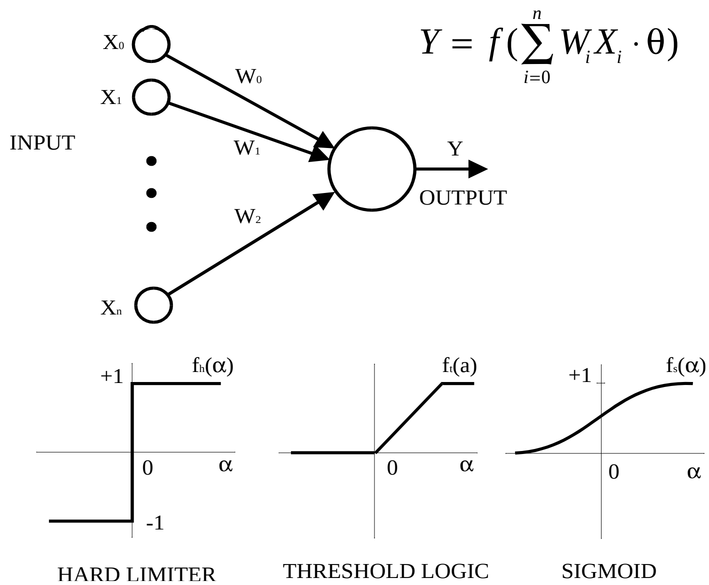
3. ANNs

3.1 Definition and structure

ANNs represent a general class of non-linear models that has been successfully applied to a variety of problems such as medical diagnostics, product selection, system control, pattern recognition, functional synthesis, and forecasting (e.g., econometrics), as well as exchange rate forecasting.

ANNs are composed of simple computational elements or nodes (Lippmann 1987). Figure 3.1 provides the simplest node, which sums N weighted inputs and conveys the outcome further. The node is characterized by an internal threshold or offset θ and by its type of specified non-linearity. Figure 3.1 illustrates three common types of non-linearities used in ANNs: hard limiters, threshold logic elements, and sigmoid. More complex nodes might even include integration or other mathematical operations.

Figure 3.1 Computational element or node with N inputs and one output (weighted sum of inputs). Three representative examples of non-linearities are shown below.

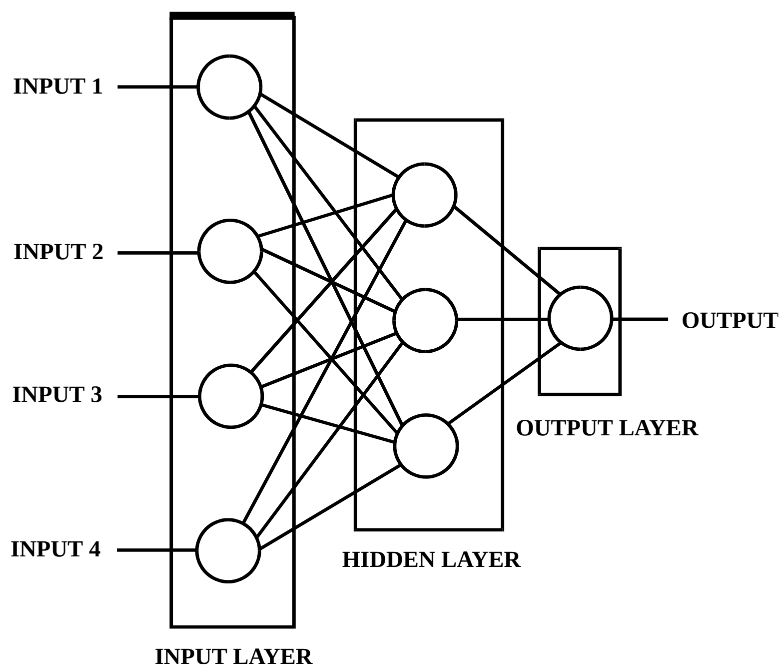


Neural network models differ in topology, node characteristics, and training or learning rules. These rules fix the initial set of weights and indicate how weights should be altered and adjusted during use to improve performance.

There are several ways to structure the neural networks. Typically, the elements are arranged in groups or layers. Fewer layers limit these networks when modelling a functional representation of data, a typical econometrics problem.

However, the development of learning algorithms has made it feasible for multi-layered networks. They are ideal for functional form determination and they are normally structured as three-layered networks (Figure 3.2).

Figure 3.2 A three-layered ANN architecture



The three layers are as follows:

- Input layer: The neural network receives its data in the input layer. The number of nodes (i.e., neurons) in this layer depends on the number of inputs to a model and each input requires one neuron. For example, in functional synthesis (this paper's scope of study), inputs are exogenous variables—that is, observations of interest.
- Hidden layer: The hidden layer lies between the input and output layers; there can be many hidden layers. They are analogous to the brain's interneurons, a place where the hidden correlations of the input and output data are captured. This allows the network to learn, adjust, and generalize from the previously learned facts (i.e., data sets) to the new input. As each input-output set is presented to the network, the internal mapping is recorded in the hidden layer. Unlike any other classical statistical methodology, this gives the system intuitive predictability and intelligence. The number of hidden layers is determined by a trade-off between network intuitive ability and efficiency. A priori, the optimal number of hidden layers is not clear. With too many hidden layers, an overcorrecting problem arises: a network is "overtrained" or

"overfitted," which prevents it from learning a general solution. On the other hand, too few layers will inhibit the learning of the input-output pattern. Typically, the number of hidden layers and nodes inside the network is determined through experimentation, and this paper follows that technique.

- Output layer: Having been trained, the network responds to new input by producing an output that represents a forecast. During training, the network collects the in-sample output values in the output layer.

Various attempts at exchange rate forecasting with ANNs are reported in Verkooijen (1996) and Plasmans, Verkooijen, and Daniels (1998), who estimated structural macroeconomic exchange rate models. In contrast, Hu et al. (1999), Zhang and Hu (1998), Kaashoek and van Dijk (1999), Kuan and Liu (1995), and Jamal and Sundar (1997) modelled the exchange rate solely as a function of its past lags. Evans (1997) forecasted the UK/DM exchange rate based on the exchange rates of several other currencies.

3.2 Learning and adapting

Neural networks develop an inner structure to solve problems. Through the training process, connection weights rearrange their values and reveal a data pattern. Thus, the fundamental feature of neural networks is that they are trained, not programmed.

A neural network learns with each new datum (an input-output combination) introduced into it during training. Every processing element responds to its input, adjusting its behaviour. The network calculates the output in accordance with the elements' transfer function. The only way to adjust to the correct response is to modify the values of the input connections. The network learns by adjusting the input weights. The equation that explains this change is called the learning law.

There are two different learning modes: supervised and unsupervised.

The supervised learning mode presents input-output data combinations to the network. Consequently, the connection weights, initially randomly distributed, adjust their values to produce output that is as close as possible to the actual output. With each subsequent cycle the error between the desired and the actual output will be lower. Eventually, the result is a minimized error between the network and actual output, as well as the internal network structure, which represents the general input-output dependence. In a one-layered network, it is easy to control each individual neuron and observe the input-output pattern. In multi-layered neural networks, supervised learning becomes difficult. It is harder to monitor and correct neurons in hidden layers. Supervised learning is frequently used for network decision, memorization, and generalization problems.

The unsupervised learning mode is independent of the external influences to adjust weights. There are no concrete data to correct the neural networks' pattern identification. Therefore, the unsupervised learning mode looks for the trend in inputs and adapts to the network function.

3.3 Backpropagation ANN

The backpropagation ANN, applied for this research, is probably the most commonly used neural network type. It is characterized by hidden layers and the generalized Delta rule for learning: If there is a difference between the actual and the desired output pattern during training, then the connection weights have to be readjusted to minimize the difference (Van Eyden 1996). Mathematically, the processing elements are modified so that they can monitor their own output. Afterwards, the actual output is compared to the model's output (I_0). The error value (E) is computed for the input pattern (X) as the difference between the actual output (y) and I_0:

$$E = I_0 - y$$

After E has been calculated, the Delta rule is applied giving the change in weights, as follows:

$$W_{new} - W_{old} = \frac{\beta \cdot E \cdot X}{X^2}$$

where: X = input
W = weight
β = the constant that measures the speed of the convergence of the weight vector

At least three layers are required: input, hidden, and output. The hidden layer is very important, since it enables the ANN to extract patterns and to generalize. Even though a hidden layer should be large, one must be careful not to deprive the network of its generalizing ability when the network starts memorizing, rather than deducing. On the other hand, a hidden layer that is too small could reduce the accuracy of recall. The connections are only feedforwarded between the adjacent layers. For the transfer function, the backpropagation ANN employs the sigmoid function (mentioned above).

The learning algorithm that controls backpropagation follows a number of steps:

1) Initialization: Initialize connection weights and neurons to small random values.
2) Data introduction: Introduce the continuous data set of inputs and actual outputs to the

backpropagation network.

3) Calculation of outputs: Calculate the outputs, and adjust the connection weights several times applying the current network error.
4) Adjustments: Adjust the activation thresholds and weights of the neurons in the hidden layer according to the Delta rule.
5) Repeat steps (2) to (4) to converge to the minimum error.

4. Model Specification

4.1 Data description

The order flow data were obtained from the Bank of Canada's unique Daily Foreign Exchange Volume Report, which is coordinated by the Bank and organized through the Canadian Foreign Exchange Committee (CFEC). Details about the trading flows (in Canadian dollars) for six major Canadian commercial banks are categorized by the type of trade (spot, forward, and futures) and transaction type (i.e., with regard to trading partner). Because this paper focuses on a short-term exchange rate forecast, spot transactions are of interest. In a spot transaction, a currency is traded for immediate delivery and payment is made within two business days of the contract entry date. Spot transactions vary, as follows:

- Commercial client transactions (CC) include all transactions with resident and non-resident non-financial customers.
- Canadian-domiciled investment transactions (CD) include all transactions with non-dealer financial institutions located in Canada.
- Foreign institution transactions (FD) include all transactions with foreign financial institutions, such as FX dealers.
- Interbank transactions (IB) include transactions with other chartered banks, credit unions, investment dealers, and trust companies in the interbank market.

Because it was unavailable prior to 1994, CD transactions are excluded as an explanatory variable in this work. However, according to D'Souza (2000), this variable is not statistically significant in exchange rate forecasting and it is not considered to be a major part of aggregate order flow.

Individual order flows (CC, FD, IB) are measured as the difference between the number of currency purchases (buyer-initiated trades) and sales (seller-initiated trades). Aggregate order flow (aggof) is the sum of individual order flows.

As noted earlier, the other variables of interest are the crude oil closing price (in U.S. dollars) deflated by the U.S. consumer price index (CPI) (Δoil) and the change in the difference between nominal 90-day commercial paper rates in Canada and the United States (Δintdiff).

The dependent variable data set comprises the logarithm of real Canada/U.S. exchange rate daily changes (Δrpfx) between January 1990 and June 2000: a total of 2,230 observations. The real exchange rate was calculated from the nominal exchange rate and CPI for the United States and Canada.

All variables are considered in first-difference terms, because the daily (weekly) change (positive or negative) prediction is of interest. Two models are considered:

ANN Model 1: $\Delta rpfx_t = f\,(\Delta intdiff_{t-j}, \Delta oil_{t-j}, aggof_{t-j}) + \varepsilon_t$, j={1, 7}; t=1,..,N.

ANN Model 2: $\Delta rpfx_t = g\,(\Delta intdiff_{t-j}, \Delta oil_{t-j}, CC_{t-j}, IB_{t-j}, FD_{t-j}) + \nu_t$, j={1, 7}; t=1,..,N.

For the purpose of ANN modelling, all data were normalized to the [-1,1] interval using the following equation:

$$x_i = \frac{X_i - X_{i,\,min}}{X_{i,\,max} - X_{i,\,min}}(h_i - l_i) + l_i$$

where: x_i = normalized value of the input or the output value

X_i = original input or output value

$X_{i,min}$ = minimum original input or output value

$X_{i,max}$ = maximum original input or output value

h_i = upper bound of the normalizing interval (in this case 1)

l_i = lower bound of the normalizing interval (in this case -1)

i=1,..,N

The ANN models were developed based on data sets of four variables (model 1) and six variables (model 2) using 2,230 observations. The model was used to forecast the daily change of the Canada/U.S. real exchange rate one day and one week into the future.

The networks trained and tested were the three-layer and four-layer backpropagation ANNs with the non-linear sigmoid and tan-sigmoid neuron activation functions in hidden layers. The number of input neurons was three for model 1 and five for model 2, while the number of hidden neurons varied between three and five for both of the models. The last layer had one linear output neuron.

The results depend strongly on the ANN architecture. More specifically, the number of hidden layers, number of neurons in the hidden layer, type of activation function, and training algorithm are principal determinants of a good prediction model. To avoid overtraining, the ANN was trained with an *early stopping* technique, where the available data set was divided into three subsets: a training set (used for gradient calculation and weights and biases updating); a validation set (when the validation error starts increasing, the training is stopped); and a testing set (used to compare real and model output, or different models).

The network development followed several steps:

- Step 1: Setting of the number of hidden layers, neurons, training algorithm (resilient backpropagation in our case), initial connection weights, and neuron biases and the activation function for each neuron.
- Step 2: Network training and validation. Sixty per cent of the time, series introduced to the ANN and the connection weights and biases values were determined within the network. The next 30 per cent was used for validation.
- Step 3: Estimation of the predicted output. The input values for the last 10 per cent of the observations were then used by the trained ANN to generate output, i.e., exchange rate forecasts.
- Step 4: Evaluation of the forecast performance of the ANN, and comparison of it to other models.
- Step 5: Steps 1–4 were repeated if the error goal was not reached.

The network training and testing was performed using the software package Matlab, v. 5.2.0, Neural Networks Toolbox, The MathWorks, Inc. (1998).

4.2 Assessment of forecast performance

In line with the Meese and Rogoff (1983) evaluation criterion, recursive estimation (or rolling regressions) will be used to evaluate the models' predictive performance. The initial estimation starts with the first 90 per cent (chronologically) of the sample *N*, or, for instance, *m* observations. That comprises training and validation sets for the ANN. The remaining 10 per cent is a testing (forecasting) set of initial size *k*; having estimated the model, *k* forecasts are generated. Subsequent (*k-1*) steps involve increasing the estimation sample (so that m increases) and shrinking the testing set (so that k decreases) by one period. In each subsequent step, (*k-s*) forecasts are estimated (*s=1,...k-1*). Finally, *k* sets of network responses (of size 1,..,k) can be compared to actual observations and other models.

This paper considers whether the ANN model can outperform linear and random walk models in terms of RMSE and the percentage of correctly predicted directions of exchange rate changes. For most of this paper, RMSE is defined as follows:

$$RMSE_{N_k} = \left[\frac{1}{N_k}\sum_{t=0}^{N_k-1}(\Delta rpf\widehat{x_{k+m-t}} - \Delta rpfx_{k+m-t})^2\right]^{\frac{1}{2}}$$

where N_k denotes the size of the out-of-sample testing set ($N_k=1,...,k$), $\Delta rpf\widehat{x_{k+m-t}}$ is the model forecast at time ($k+m-t$), and $\Delta rpfx_{k+m-t}$ is the actual exchange rate change. The ratio of data allocated to training, validation, and testing was maintained at 6:3:1 throughout the recursive experiment.

In addition to RMSE, the percentage of correctly predicted signs (PERC) of the forecasted variable $\Delta rpfx_t$ is considered; this is the total number of correctly forecasted positive and negative movements, defined as:

PERC(N_k) = (number of positive correct responses + number of negative correct responses)/N_k

5. Empirical Results

This section assesses the forecasting performance of a range of exchange rate models. Generally, the random walk model performs better than any traditional linear macroeconomic model that excludes microstructure variables; therefore, it can be viewed as the benchmark model. The following models are considered first:

Random walk model (RW):

$rpfx_t = \alpha_0 + rpfx_{t-j} + \gamma_t$, j={1, 7}; t=1,..,N.

Linear model 1:

$\Delta rpfx_t = \gamma_0 + \gamma_1 \Delta intdiff_{t-j} + \gamma_2\, \Delta oil_{t-j} + \gamma_3\, aggof_{t-j} + \varepsilon_t$, j={1, 7}; t=1,..,N.

Linear model 2:

$\Delta rpfx_t = \beta_0+\beta_1\Delta intdiff_{t-j} + \beta_2 \Delta oil_{t-j} + \beta_3 CC_{t-j} +\beta_4 IB_{t-j} +\beta_5 FD_{t-j} + \nu_t$, j={1, 7}; t=1,..,N.

Table 5.1 presents linear regression estimation results for these models based on the first 2,005 observations (initial estimation set). The impact of interest rate change is more significant for lower frequency models (j=7), while the estimator of oil price change is more significant for one-day-ahead forecasting, Also, order flows are more important for higher frequency forecasting. Even though it is very small, as expected, the R^2 increased when individual order flows were taken into account.[7]

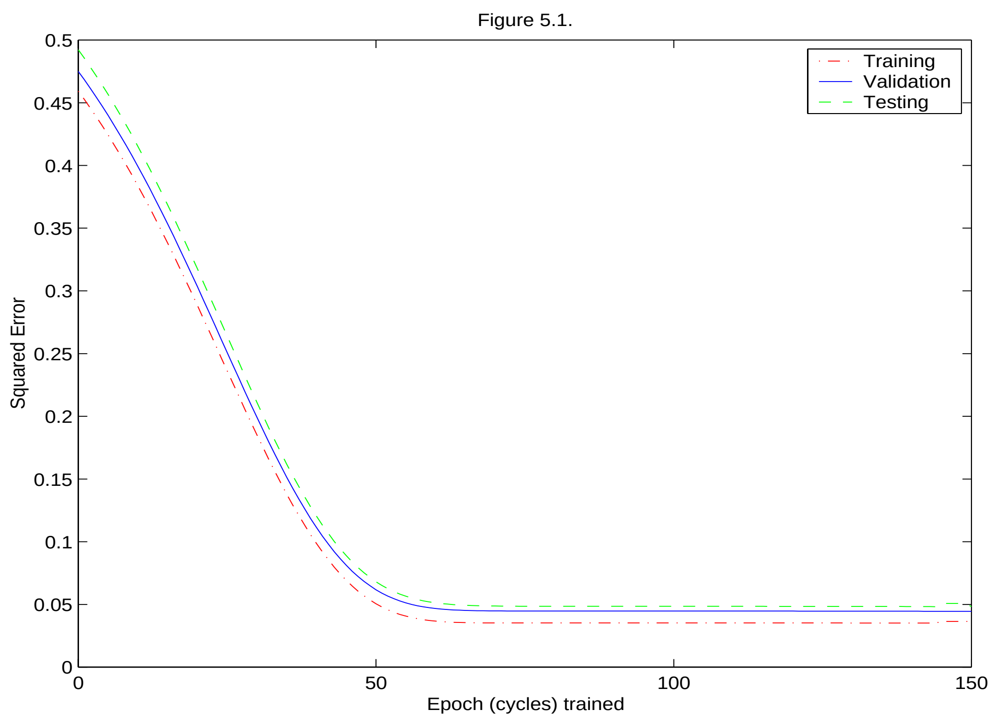
Table 5.1 Estimation Results For Linear Models

Estimates (standard error)	Model			
	Linear Model 1 (j=1)	Linear Model 1 (j=7)	Linear Model 2 (j=1)	Linear Model 2 (j=7)
γ_0 (exp 10^{-5})	8.09 (2.93e-05)	36.53 (7.08e-05)		
β_0 (exp 10^{-5})			8.72 (2.95e-05)	36.35 (7.15e-05)
$\Delta intdiff_{t-j}$ (exp 10^{-4})	-1.13 (0.00025)	-6.76 (0.00026)	-1.89 (0.00025)	-6.63 (0.00026)
Δoil_{t-j}	-0.0090 (0.0065)	-0.003 (0.0073)	-0.0087 (0.0065)	-0.003 (0.0073)
$aggof_{t-j}$ (exp 10^{-7})	-1.35 (8.96e-08)	-3.74 (2.16e-07)		
CC_{t-j} (exp 10^{-7})			1.33 (1.28e-07)	-4.55 (3.11e-07)
IB_{t-j} (exp 10^{-7})			-1.015 (1.69e-07)	-2.66 (4.07e-07)
FD_{t-j} (exp 10^{-7})			-3.86 (8.98e-08)	-3.78 (2.16e-07)
R^2	0.0021	0.005	0.0049	0.0055

7. This paper uses daily data over a ten-year period (as opposed to the four-month span used by Lyons and Evans 1999); therefore, the linear "microstructure" model's R^2 is significantly lower.

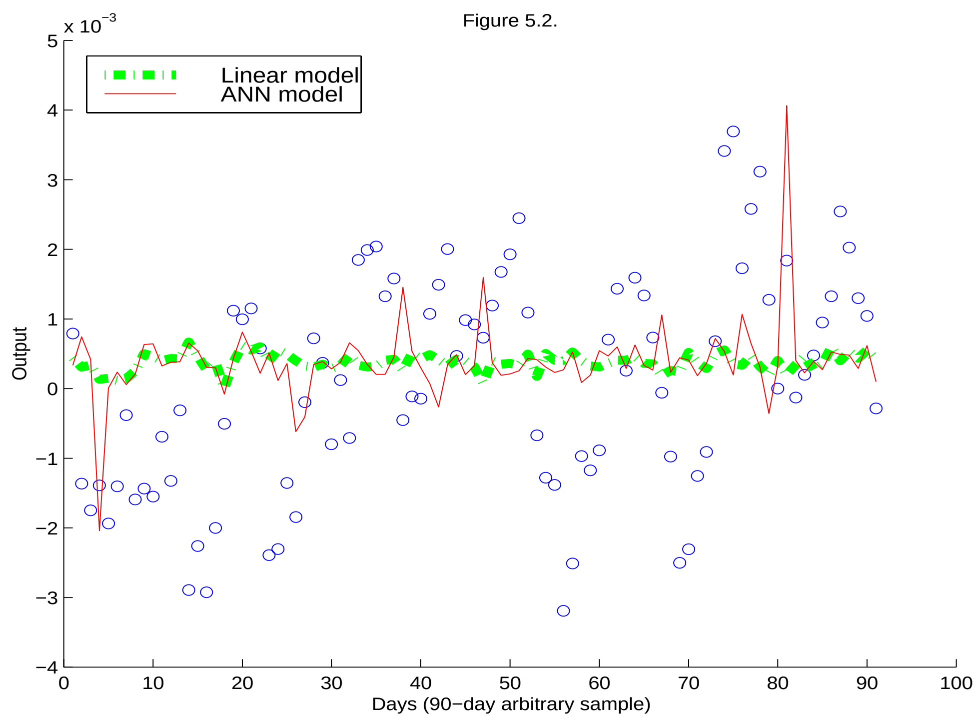
Two non-linear models, as described in Section 4 (ANN models 1 and 2), were estimated by feedforward backpropagation ANN. Figure 5.1 shows the errors related to the training set, the validation set, and the testing set. As expected, all three errors decline during the learning process. Overtraining was prevented by stopping the training process when the validation set error started to increase.

Figure 5.1 ANN errors: training, validation, and testing



Full sample estimation of 2,230 observations was used to compare the ANN and linear model's performance. Figure 5.2 shows that the linear model forecasts in a linear fashion (for an arbitrary 90-day period), whereas the ANN forecasts more in keeping with the pattern of actual exchange rate changes.

Figure 5.2 Linear and ANN model exchange rate forecasts. Actual values are denoted by circles.



After the initial estimation of the models in the first 2,005 observations, a set of out-of-sample forecasts was used to generate RMSEs. Each recursive re-estimation added 10 observations, so that 18 RMSEs were calculated on out-of-sample data sets ranging in size from 225 to 55 observations. This led to the selection of an ANN model 1 and ANN model 2 for one-day-ahead (j=1) and for one-week-ahead (j=7) forecasts of exchange rate changes, which were compared to linear models 1 and 2 and the random walk model.

Figures 5.3, 5.4, 5.5, and 5.6 show that the ANN can produce promising short-run forecasts, since the RMSE for the ANN model for a given forecasting horizon is equal to or below both of the competing models.

Figure 5.3 RMSE for ANN model 1, linear model 1, and random walk (j=1)

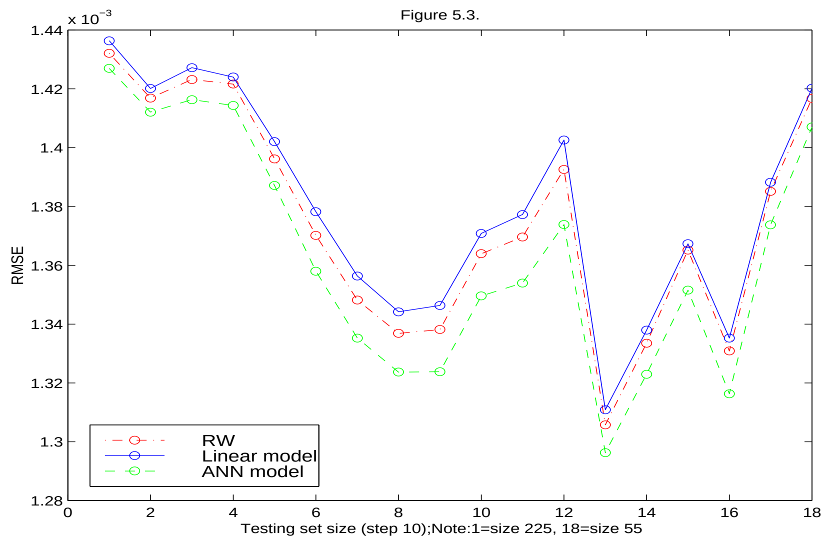


Figure 5.4 RMSE for ANN model 2, linear model 2, and random walk (j=1)

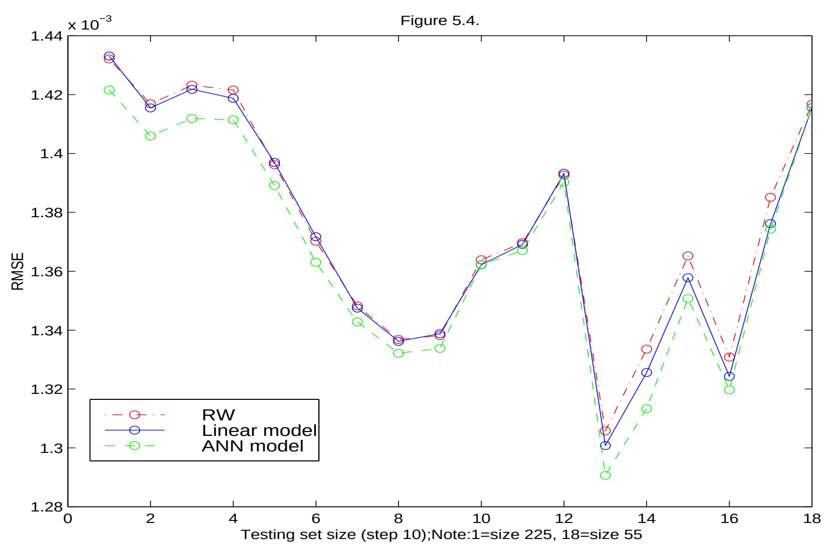


Figure 5.5 RMSE for ANN model 1, linear model 1, and random walk (j=7)

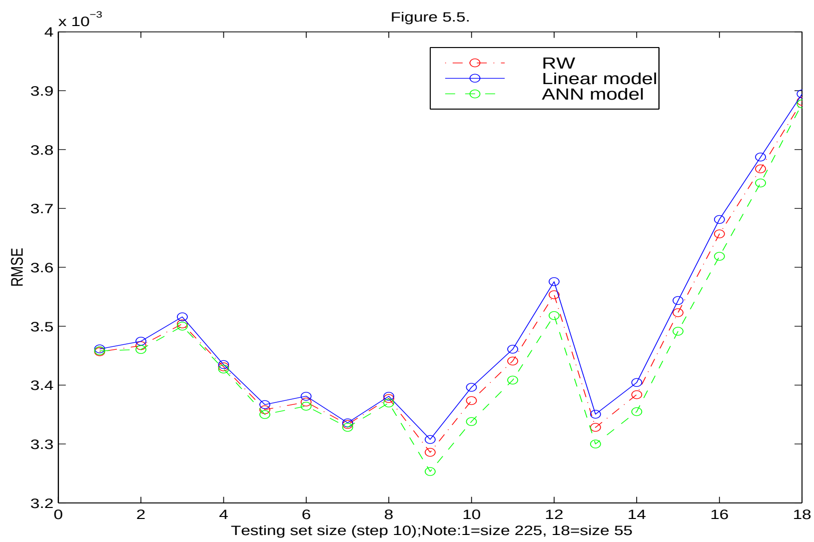
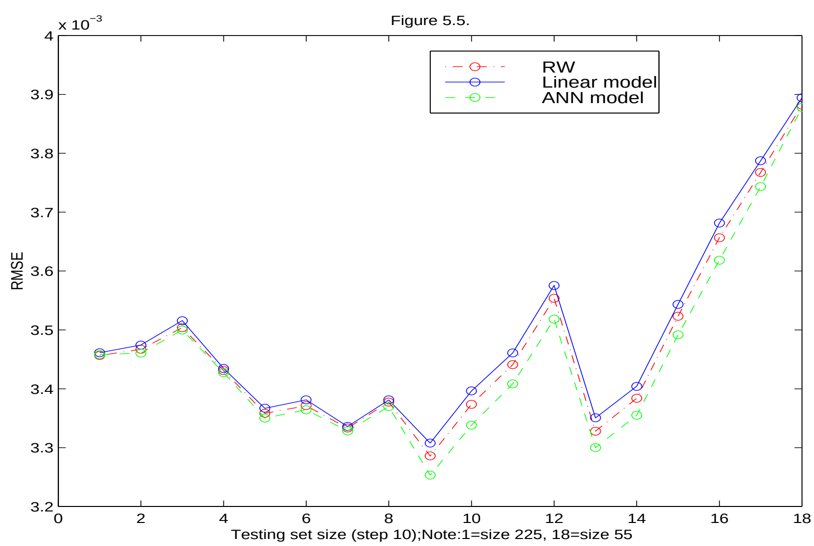


Figure 5.6 RMSE for ANN model 2, linear model 2, and random walk(j=7)



Tables 5.2 (for j=1) and 5.3 (for j=7) list the RMSE statistics illustrated in these figures.

Table 5.2 RMSE (exp 10^{-3}) for ANN, linear, and random walk models (j=1)

Sample size	Model				
	Random Walk (j=1)	Linear Model 1 (j=1)	ANN Model 1 (j=1)	Linear Model 2 (j=1)	ANN Model 2 (j=1)
225	1.4321	1.4364	1.4270	1.4331	1.4216
215	1.4168	1.4201	1.4121	1.4155	1.4060
205	1.4232	1.4272	1.4163	1.4218	1.4119
195	1.4216	1.4240	1.4143	1.4188	1.4114
185	1.3962	1.4020	1.3871	1.3969	1.3892
175	1.3702	1.3782	1.3580	1.3717	1.3631
165	1.3482	1.3564	1.3353	1.3475	1.3428
155	1.3368	1.3442	1.3237	1.3362	1.3322
145	1.3381	1.3464	1.3238	1.3388	1.3337
135	1.3639	1.3708	1.3496	1.3622	1.3622
125	1.3696	1.3773	1.3540	1.3691	1.3671
115	1.3926	1.4026	1.3739	1.3933	1.3903
105	1.3057	1.3109	1.2963	1.3008	1.2907
95	1.3335	1.3380	1.3230	1.3256	1.3133
85	1.3652	1.3673	1.3515	1.3578	1.3508
75	1.3308	1.3353	1.3163	1.3243	1.3198
65	1.3851	1.3882	1.3738	1.3761	1.3743
55	1.4168	1.4202	1.4070	1.4156	1.4155

Table 5.3 RMSE (exp 10^{-2}) for ANN, linear, and random walk models (j=7)

Sample size	Model				
	Random Walk (j=7)	Linear Model 1 (j=7)	ANN Model 1 (j=7)	Linear Model 2 (j=7)	ANN Model 2 (j=7)
225	0.3457	0.3461	0.3458	0.346	0.3448
215	0.3467	0.3474	0.3461	0.3473	0.3462
205	0.3504	0.3516	0.35	0.3515	0.3502
195	0.3431	0.3435	0.3427	0.3437	0.3428
185	0.3358	0.3367	0.335	0.3368	0.3358
175	0.3371	0.3381	0.3364	0.3382	0.3365
165	0.3333	0.3336	0.3328	0.3338	0.3331
155	0.3377	0.3381	0.337	0.3382	0.3373
145	0.3286	0.3308	0.3253	0.3307	0.3278
135	0.3374	0.3396	0.3338	0.3396	0.3366
125	0.3441	0.3461	0.3408	0.3462	0.3438
115	0.3554	0.3576	0.3518	0.3577	0.3553
105	0.3328	0.3351	0.33	0.3349	0.3326
95	0.3384	0.3404	0.3355	0.3404	0.3384
85	0.3523	0.3544	0.3491	0.3544	0.3521
75	0.3657	0.3682	0.3619	0.3683	0.3656
65	0.3768	0.3787	0.3743	0.3790	0.3764
55	0.3882	0.3894	0.3878	0.3895	0.3868

The experiments show that the ANN model forecasts one-day and seven-day-ahead exchange rate changes better than the linear and random walk models. Nevertheless, the primary indicator of good forecasting power is not necessarily RMSE, but the percentage of correctly forecasted directions of real exchange rate fluctuations. In this case, the estimation involves very small values (exp 10^{-3}) that might result in small RMSEs. In turn, the presence of small RMSEs is not a guarantee that the prediction is accurate, and caution is required when interpreting the estimation results.

As noted above, the percentage of correctly forecasted exchange rate direction changes or good hits (PERC) is also considered. Recursive regression for horizons between 5 and 225 observations (step 5) reveals the superiority of the ANN model.[8] ANN model 1 (2) correctly predicted, on average, 60.14 per cent (61.81 per cent) of the direction of daily exchange rate movements, while linear model 1 (2) correctly predicted 57.18 per cent (58.75 per cent) of such changes, and the random walk model predicted 54.88 per cent. One-week-ahead forecasts yield worse results for ANN model 1 and linear model 1 against random walk for j=7, but ANN model 2 has the best results. Also, the predictive power of both non-random walk models is lower. Table 5.4 compares all the models used in terms of the second comparison criterion.

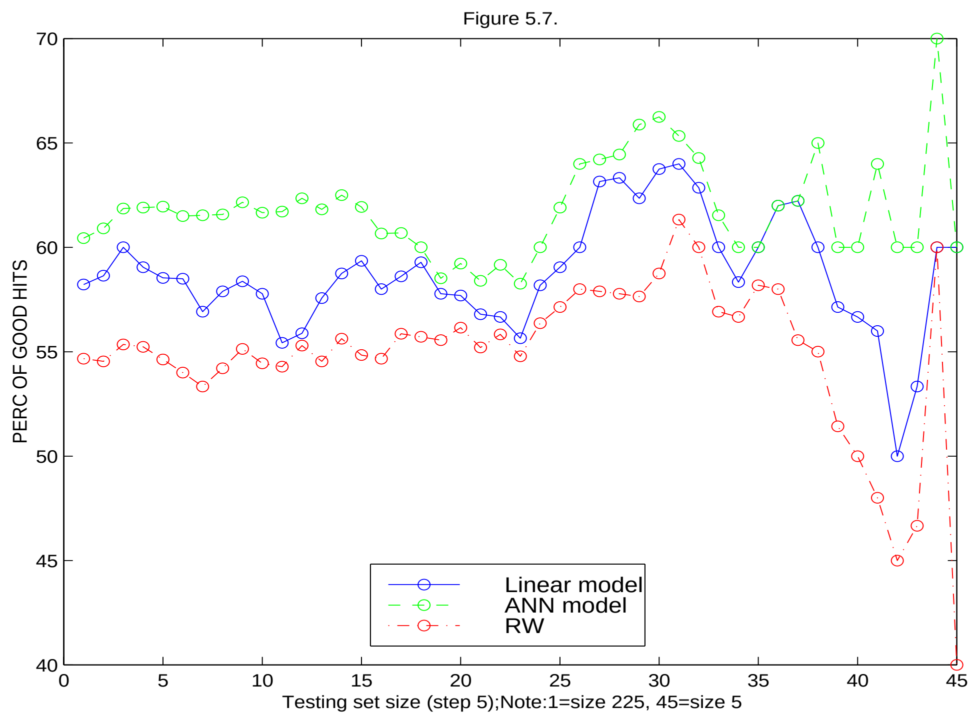
Table 5.4 The average percentages of correctly predicted signs for linear models 1 and 2 (LM 1 and LM 2), ANN models 1 and 2 (ANN 1 and ANN 2), and the random walk model. One-day (j=1) and one-week (j=7) forecasts are considered.

AVERAGE PERC (%)	Model				
	Random Walk	LM 1	ANN 1	LM 2	ANN 2
j=1	54.88	57.18	60.14	58.75	61.81
j=7	56.26	54.9	56.15	55.28	58.04

8. Step 5 is used instead of step 10 to impose a more demanding setting for ANN models.

Figure 5.7 illustrates the one-day results summarized in Table 5.4. The results show that the ANN models dominate in predicting the direction of exchange rate changes one day ahead.

Figure 5.7 Example: recursive estimation PERC comparison for ANN model 2, linear model 2, and random walk model (1-day forecast)



Next, out-of-sample, one-step-ahead forecasts are considered. More precisely, ANN model 1 (2) is initially estimated for the first 2,006 observations. The forecast errors for the remaining 225 observations (a testing set) are calculated by extending the estimation set by one and recalculating the forecast errors until the whole testing set is exhausted. This differs from the preceding forecast experiment in that the earlier experiment did not re-estimate the model up to t-1 to forecast the exchange rate at t. RMSEs and PERCs for the one-day-ahead forecasts are listed in Table 5.5. The striking result here is that ANN 2 correctly predicts almost 72 per cent of the directions of future exchange rate changes, while the random walk model stays at about a 55 per cent accuracy.

Table 5.5 PERC and RMSE (exp 10^{-3}) statistics for the recursive estimation over the whole testing set (k=225). ANN models 1 and 2 (ANN 1 and ANN 2) and the random walk (RW) model for one-day-ahead (j=1) forecasts are considered.

	Model		
	RW	ANN 1	ANN 2
PERC (%):	54.88	67.56	71.56
RMSE:	1.4321	1.4155	1.3988

To determine the percentage of correctly predicted changes (or good hits) that relates to positive changes, the following statistic was constructed for the initial testing sample size (k=225):

PERC(POS) = (number of positive correct responses/number of sample positive movements)

Similarly, for negative good hits another statistic was calculated:

PERC(NEG) = (number of negative correct responses/number of sample negative movements)

The term “positive changes” refers to values above the mean of estimation sample changes, while “negative changes” are values below the mean value. This corrects for the fact that there is a significantly greater number of positive changes in this sample. Taking zero as a mean value would affect the reliability of the criterion, since there were mostly positive changes in the sample.

According to Table 5.6, the ANN models forecast positive and negative changes roughly equally well. In comparison, failing to correct for the positive mean change would lead to the erroneous conclusion that the model predicts positive changes much better than negative changes.

Table 5.6 PERC(POS) and PERC(NEG) for ANN models 1 and 2 (ANN 1 and ANN 2). One-day (j=1) and one-week (j=7) forecasts are considered (k=225). Percentages without normalization are in parentheses.

	Model	
	ANN 1	ANN 2
PERC(POS) (%):		
j=1	42.98 (89.43)	38.02 (80.49)
j=7	57.02 (98.39)	53.04 (98.39)
PERC(NEG) (%):		
j=1	48.08 (2.05)	67.31 (36.27)
j=7	41.44 (0.99)	56.36 (2.97)

6. Conclusions and Further Research

This paper combines two new approaches—artificial neural networks and market microstructure—to exchange rate determination to explain very short-run exchange rate fluctuations. A variable from the field of microstructure, order flow (aggregate and its components), is included in a set of macroeconomic variables (interest rate and crude oil price) to explain Canada/U.S. dollar exchange rate movements.

The results show that the ANN model never performs worse than the linear model, and always better than the random walk model. This result is not surprising, since ANN is able to model any non-linear as well as linear functional dependencies. Thus, appropriately selected ANN models dominated linear models and produced better out-of-sample forecasts.

Two criteria are applied to evaluate model performance: RMSE and the ability to correctly predict the direction of the exchange rate movements. The ANN is consistently better in terms of RMSE than random walk and linear models for the various out-of-sample experiments. Moreover, ANN performs on average at least 3 per cent better than other models in its percentage of correctly predicted signs. This is true for both of the forecasting horizons. As expected, more accurate forecasts are generated for the shorter forecasting window, but they are still superior to the random walk model. Recursive one-step-ahead forecasts lead to a considerable improvement in PERC compared to the random walk and linear models.

The results indicate that both macroeconomic and microeconomic variables are useful to forecast high-frequency exchange rate changes. The inclusion of other microeconomic and macroeconomic variables could improve these findings. Further, including only significant lags of independent variables could give more accurate forecasts. The significance of each lag should be tested by ANN models. As well, more complete, longer time series and even higher-frequency data might be fruitful. The power of this approach can be tested on other currencies. Finally, the ANN developed for this research can be modified in terms of ANN type, topology, and learning rule. Connecting (on an adequate basis) the ANN with the statistical techniques, genetic algorithms (Goldberg 1989), fuzzy logic (Cox 1992), and expert systems (Watkins 1993) is a research direction where high payoffs can be expected.

References

Amano, R.A. and S. van Norden. 1995. “Terms of Trade and Real Exchange Rates: The Canadian Evidence.” *Journal of International Money and Finance* 14 (1): 83–104.

———. 1998. “Exchange Rates and Oil Prices.” *Review of International Economics* 6 (4): 683–94.

Baillie, R. and P. McMahon. 1989. *The Foreign Exchange Market: Theory and Econometric Evidence.* New York: Cambridge University Press.

Covrig, V. and M. Melvin. 1998. “Asymmetric Information and Price Discovery in the FX Market: Does Tokyo Know More about the Yen?” Arizona State University. Typescript.

Cox, E. 1992. “Fuzzy Fundamentals.” *IEEE Spectrum* 29 (10): 58–61.

D’Souza, C. 2000. “Order-Flow and Exchange Rate Forecasting.” Bank of Canada mimeo, 1-30.

Evans, O.V.D. 1997. “Short-Term Currency Forecasting using Neural Networks.” *ICL Systems Journal* 11 (2): 1–17.

Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley Publishing Co.

Hsieh, D. 1989. “Testing for Nonlinear Dependence in Daily Foreign Exchange Rate Changes.” *Journal of Business* 62: 329–68.

Hu, M.Y., G. Zhang, C.X. Jiang, and B.E. Patuwo. 1999. “A Cross-Validation Analysis of Neural Network Out-of-Sample Performance in Exchange Rate Forecasting.” *Decision Sciences* 30 (1): 197–216.

Jamal, A.M.M. and C. Sundar. 1997. “Modeling Exchange Rate Changes with Neural Networks.” *Journal of Applied Business Research* 14 (1): 1–5.

Kaashoek, J.F. and H.K. van Dijk. 1999. “Neural Networks Analysis of Varying Trends in Real Exchange Rates.” Erasmus University Rotterdam, Econometric Institute Report 143, 1–19.

Kuan, C.-M. and T. Liu. 1995. “Forecasting Exchange Rates using Feedforward and Recurrent Neural Networks.” *Journal of Applied Econometrics* 10 (4): 347–64.

Lippmann, R.P. 1987. “An Introduction to Computing with Neural Nets.” *IEEE ASSP Magazine*, 4–22.

Lyons, K.R. and M.D.D. Evans. 1999. “Order Flow and Exchange Rate Dynamics.” NBER Working Paper No. W7317, 1–46.

Meese, R.A. and K. Rogoff. 1983. “Empirical Exchange Rate Model of the Seventies: Do They Fit Out of Sample?” *Journal of International Economics* 14: 3–24.

Meese, R.A. and A.K. Rose. 1990. “Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation.” *The American Economic Review* 80 (2): 678–91.

———. 1991. “An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination.” *Review of Economic Studies* 58 (3): 603–19.

Plasmans, J., W. Verkooijen, and H. Daniels. 1998. “Estimating Structural Exchange Rate Models by Artificial Neural Networks.” *Applied Financial Economics* 8: 541–51.

Van Eyden, R.J. 1996. *The Application of Neural Networks in the Forecasting of Share Prices*. Haymarket, VA: Finance & Technology Publishing.

Verkooijen, W. 1996. “A Neural Network Approach to Long-Run Exchange Rate Prediction.” *Computational Economics* 9: 51–65.

Watkins, P. and L.B. Eliot. 1993. *Expert Systems in Business & Finance: Issues & Applications*. Chichester: John Wiley & Sons Canada, Limited.

Yao, J. 1997. “Spread Components and Dealer Profits in the Interbank Foreign Exchange Market.” New York University, Salomon Center, Working Paper: S/98/04, 1–54.

Zhang, G. and M.Y. Hu. 1998. “Neural Network Forecasting of the British Pound/US Dollar Exchange Rate.” *International Journal of Management Science* 26 (4): 495–506.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*****.**

2000

2000-22	Une analyse empirique du lien entre la productivité et le taux de change réel Canada-É-U	D. Dupuis et D. Tessier
2000-21	Les effets réels du cours des actions sur la consommation	L. Pichette
2000-20	Steps in Applying Extreme Value Theory to Finance: A Review	Y. Bensalah
2000-19	Le modèle USM d'analyse et de projection de l'économie américaine	R. Lalonde
2000-18	Inflation and the Tax System in Canada: An Exploratory Partial-Equilibrium Analysis	B. O'Reilly and M. Levac
2000-17	A Practical Guide to Swap Curve Construction	U. Ron
2000-16	Volatility Transmission Between Foreign Exchange and Money Markets	S.K. Ebrahim
2000-15	Private Capital Flows, Financial Development, and Economic Growth in Developing Countries	J.N. Bailliu
2000-14	Employment Effects of Nominal-Wage Rigidity: An Examination Using Wage-Settlements Data	U.A. Faruqui
2000-13	Fractional Cointegration and the Demand for M1	G. Tkacz
2000-12	Price Stickiness, Inflation, and Output Dynamics: A Cross-Country Analysis	H. Khan
2000-11	Identifying Policy-makers' Objectives: An Application to the Bank of Canada	N. Rowe and J. Yetman
2000-10	Probing Potential Output: Monetary Policy, Credibility, and Optimal Learning under Uncertainty	J. Yetman
2000-9	Modelling Risk Premiums in Equity and Foreign Exchange Markets	R. Garcia and M. Kichian
2000-8	Testing the Pricing-to-Market Hypothesis: Case of the Transportation Equipment Industry	L. Khalaf and M. Kichian
2000-7	Non-Parametric and Neural Network Models of Inflation Changes	G. Tkacz
2000-6	Some Explorations, Using Canadian Data, of the *S*-Variable in Akerlof, Dickens, and Perry (1996)	S. Hogan and L. Pichette

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à:

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9

E-mail / Adresse électronique: publications@bankofcanada.ca
WWW: http://www.bankofcanada.ca/